ESD-TR-88-158

(2)

Technical Report
803

On Random Correlation Matrices

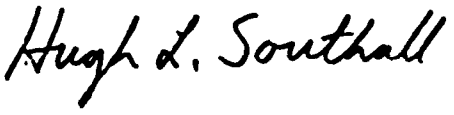
R.B. Holmes

28 October 1988

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

**Prepared for the Department of the Air Force
under Electronic Systems Division Contract F19628-85-C-0002.**

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract F19628-85-C-0002.



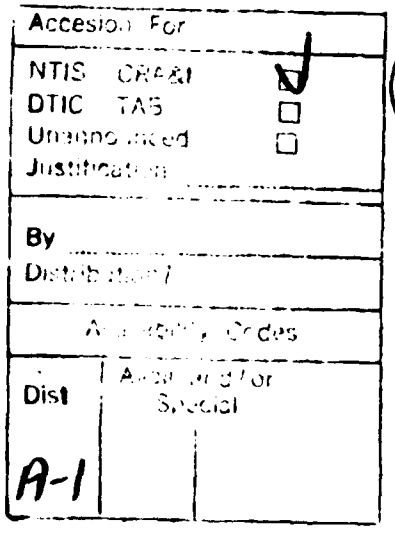
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

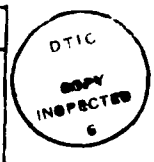
ON RANDOM CORRELATION MATRICES

R.B. HOLMES
Group 32

TECHNICAL REPORT 803

28 OCTOBER 1988





LEXINGTON MASSACHUSETTS

ABSTRACT

This report contains a detailed study of random correlation matrices, including algebraic, statistical, and historical background. Such matrices are of particular interest because they serve to model 'average signals' for simulation testing of signal processing algorithms. The latter half of this report extensively discusses the statistical behavior of spectral functions of the two major types of random correlation matrices, from both theoretical and empirical aspects. Our emphasis then, is on eigenvalue distribution and condition number behavior. Actual application to algorithm testing will be described in a subsequent report.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ON RANDOM CORRELATION MATRICES

1. INTRODUCTION

This report is based on a study of the relative efficacy of certain (group-theoretic) data transforms for various canonical signal processing tasks. Two such tasks are, in particular, data compression and decorrelation. For a given data transform, realized as a unitary matrix U, the extent of such activity can be measured from the transformed data covariance matrix. Thus if a data vector x has covariance C, its transform Ux has covariance D = UCU*, and the data compression (respectively decorrelation) efficiency of the transform U can be assessed by examination of the diagonal (respectively off-diagonal) entries of D.

In order to make a serious statistical study of the efficiency of group transforms and filters for the various signal processing tasks, it is necessary to have an assortment of standardized signal models. These fall into two classes: parametric models and 'purely random' models. The former determine after sampling structured covariance matrices with entries having a simple dependence on a few parameters. The simplest and most familiar example of this model type is the first-order Markov or auto-regressive signal model, from which N samples generate the covariance matrix $[\rho^{|i-j|}]$, where $0 < |\rho| < 1$, and $1 \leqslant i, j \leqslant N$. It is somewhat less clear, *a priori*, what a 'purely random' covariance structure might be. The object of the following sections is to clarify and discuss the term 'purely random'. Speaking intuitively for the moment, this term must be precisely defined if we are to seriously simulate the action of the various transforms, and to eventually say that one or another of them, for fixed data dimension, is superior in the performance of a particular task 'on the average.'

1.1 Definitions

In the background we have an N-dimensional real or complex-valued second order random vector x. We will usually assume that x has 0-mean: $E(x) = \theta$, the zero vector. The covariance matrix of x is the $N \times N$ matrix C_x defined by

$$C_x = [E(x_i \bar{x}_j)] \quad .$$

Such matrices are characterized as being Hermitian and positive semidefinite. We will in fact always assume that C_x is actually positive definite, so as to eliminate degenerate probability density functions. Hence the eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$ of C_x are all positive; they constitute the spectrum $\sigma(C_x)$ of C_x, and their relative size will always be indicated by subscript:
$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_N > 0$.

We recall the statistical significance of these eigenvalues: letting $\{\phi_1, \ldots, \phi_N\}$ be the orthonormal set of eigenvectors corresponding to $\lambda_1, \ldots, \lambda_N$, we have

(a) $\lambda_i = \text{var}(< x, \phi_i >)$

(b) $\text{tr}(C_x) = \lambda_1 + \ldots + \lambda_N = E(\|x\|^2)$

(c) $\lambda_{m+1} + \ldots + \lambda_N = \min_{S_m} E(d(x, S_m)^2)$, $m = 1, \ldots, N-1$.

The assertion here is that λ_i is the variance of the ith principal component of x; these random variables occur as the coefficients in the expansion of x in the (Karhounen-Loeve) basis $\phi_1, \ldots, \phi_N$. Statement (b) is a special case of (c) (take m = 0, there). The final assertion is that the best mean square approximation to x by m-dimensional subspaces, S_m, occurs when S_m is spanned by $\phi_1, \ldots, \phi_m$, with error as the indicated function of the eigenvalues. For applications of these and related formulas to multivariate statistics, pattern recognition, and signal processing (estimating x from noisy observations), see, respectively References 1, 2, and 3.

From now on we will make a slight specialization by assuming that all components of the random vector x have the same variance, which we take to be unity. It follows that the diagonal of C_x consists of ones, $\mathrm{tr}(C_x) = N$, and the modulus of each off-diagonal entry c_{ij} satisfies $|c_{ij}| < 1$. These entries are, in fact, the correlation coefficients of the i and j components of x. Any such matrix is called a correlation matrix, and will be our primary object of study. Bounds and estimates for various quantities associated with such matrices are reviewed in Section 3. Here we note that if C is any $N \times N$ correlation matrix, then $1 \leqslant \|C\| = \lambda_1 \leqslant \lambda_1 + \ldots + \lambda_N = N$, so that the set $\Gamma(N)$ of all such C is a compact convex subset of the $N(N+1)/2$ -dimensional real space of $N \times N$ Hermitian matrices. (If the scalars are complex, this latter space is of real dimension N^2.)

In general, it is difficult to tell by inspection whether a given symmetric or Hermitian matrix C with diagonal entries equal to one is positive definite, and hence a correlation matrix. Several nonlinear inequalities involving the off-diagonal entries must be satisfied; these correspond to the positivity of the leading principal minors of C. For example, given real numbers b, c, d in (-1,1), the matrix

$$\begin{bmatrix} 1 & b & c \\ b & 1 & d \\ c & d & 1 \end{bmatrix}$$

is a correlation matrix if and only if

$$b^2 + c^2 + d^2 < 1 + 2bcd.$$

Two simple sufficient conditions for positive definiteness are available, however:

(a) C is diagonally dominant, so that the Gershgorin theorem can be applied, and

(b) C can be partitioned as

$$C = \begin{bmatrix} I_1 & F \\ F^* & I_2 \end{bmatrix}$$

where the I's are identity matrices, and F is a matrix whose (spectral) norm is less than one.

1.2 Notions of Randomness

We now want to address the question of randomly selecting a correlation matrix of some fixed size. Our particular interest in this question has already been indicated in the introductory remarks above, and further motivation will be provided in the next section; in general we may say simply that a satisfactory answer to this question will permit generation of random test problems for a variety of statistical methods.

Roughly speaking, any method of generating random correlation matrices will begin by generating some number of pseudorandom variates uniformly distributed on the unit interval, and then performing certain deterministic mathematical steps to arrive at a correlation matrix. Four possible such methods will be described below, and two will be discussed at some length. But we have to acknowledge at the outset that no method is completely satisfactory. This is due to the lack of structure of the set $\Gamma(N)$ on the one hand, and to the presence of structure in the individual members of $\Gamma(N)$ on the other hand. That is, each element C of $\Gamma(N)$ has associated with it, as a matrix, entries, eigenvalues, and functions of these, such as norm, condition number, etc., all of which become random variables with their own distributions which naturally depend on the manner in which C was produced. But the set $\Gamma(N)$ does not carry a natural invariant measure. This deficiency may be contrasted with the cases of the orthogonal or unitary groups, which carry a canonical (unimodular) Haar measure. Nor is $\Gamma(N)$ a homogeneous space, such as a sphere, on which a transformation group acts and leaves invariant some measure. Thus, while such phrases as 'random orthogonal matrix' and 'uniform distribution over the unit sphere S^{N-1}, have a clear conceptual meaning, and indeed there exist successful numerical procedures for generating such variates (see in particular References 4 and 5 for the former case), the situation remains murky for correlation matrices.

As a brief aside we offer two remarks. First, the topics of approximating and efficiently sampling from the uniform (Haar) distribution on finite or compact groups persist under current investigation. In addition to the references just given for the case of the orthogonal group, see recent articles by L. Takacs[6] for finite groups, and by P. Diaconis and various coauthors (Reference 7 and its references) for an assortment of groups and applications. The basic approximation result, that the distributions of the successive terms in a random walk on a compact group converge vaguely to normalized Haar measure provided that the support of the common distribution of the individual terms is sufficiently diffuse, goes back at least to Grenander.[8] The condition on the support of the distribution can also be rephrased as a spectral property of its (operator-valued) Fourier transform. Second, although as noted above, $\Gamma(N)$ is not a homogeneous space; the cone P(N) of all positive definite $N \times N$ Hermitian matrices is such a space. Namely, it is acted on by the general or special linear groups according to the rule

$$A \rightarrow T A T^*$$

for $A \in P(N)$ and T nonsingular. It follows from general theory, including the fact that these linear groups are unimodular, that there is an invariant measure on P(N).[9]

Returning now to the matter of random correlation matrices we indicate four possible methods of generation; only the last two will be discussed in any detail, beginning in the next section.

Method 1: Direct Acceptance — Rejection

Here one must obtain symbolically the leading principal minors of the general symmetric $N \times N$ matrix with unit diagonal. This is possible for moderate size N via a computer algebra program. Requiring these minors to be positive then constitutes a set of $N - 2$ nonlinear constraints on the $N(N - 1)/2$ off-diagonal entries. A set $C_{12}, \ldots, C_{1N}, C_{23}, \ldots, C_{2N}, \ldots, C_{N-1,N}$ of pseudorandom deviates uniformly distributed on $(-1,1)$, or in the open unit disc, is generated, and tested to see if the constraints are satisfied. If the constraints are satisfied, a correlation matrix C is obtained; if not, a new set of uniform deviates is generated, etc. Clearly this method is at best feasible for rather small values of N, say $N \leqslant 8$. To the author's knowledge, the distributional aspects of the spectral features of the resulting matrices are unknown.

Method 2: Perturbation about a Mean

This method is discussed by Marsaglia and Olkin,[10] which is generally the most current source of information about our subject. However, for our purposes, there is no reason to have in mind any *a priori* mean value.

Method 3: Random Spectrum

As we know, the spectrum of an $N \times N$ correlation matrix consists of N positive numbers (not necessarily distinct) that sum to N. As will be recalled in the next section, every such set of N numbers occurs, so that the possible spectra fill out a simplex in real N-space. Since it is numerically feasible to generate pseudorandom uniform samples from this simplex, we can, by a succession of suitably chosen orthogonal or unitary transforms, arrive at a random correlation matrix. An automated procedure for doing this latter task is commercially available in the IMSL subroutine GGCOR. Statistical aspects of this method will be discussed at some length in Section 4.

Method 4: Random Gram Matrix

As is well known, every real positive definite matrix A has a Cholesky factorization

$$A = TT^*$$

where T is a uniquely defined lower triangular matrix with positive diagonal entries. Let the rows of T be denoted $t_1, \ldots, t_N$. Then

$$a_{ij} = \langle t_i, t_j \rangle$$

and so A can be considered a Gram matrix defined by the vectors $\{t_1, \ldots, t_N\}$. If also A is a correlation matrix, then each vector t_i must have length 1. Consequently, any procedure for generating pseudorandom unit vectors, with any distribution, will result in a random correlation

matrix of Gram type. These vectors may or may not end in zeros, as in the Cholesky factorization, but naturally we do less work if they do. This method is the most efficient of the general methods 1, 3, and 4; some of its statistical aspects will be discussed in Section 5 (see also Reference 10 again).

The method we might eventually choose for a particular application will depend on the nature of the application and just which aspect of the random correlation matrices we wish to have an unambiguous uniform distribution.

1.3 Background and Motivation

In terms of the preceding introductory material, and prior to the more technical developments of the remaining sections, we will briefly review some of the relevant statistical literature. Specifically, we will comment on the contents of four articles, References 11, 12, 13, and 10, listed in chronological order.

Chalmers (1975, Reference 11) presents an algorithm which produces correlation matrices with a common spectrum. His motivation is the study of strongly structured data, that is , random vectors whose first two or three principal components explain much of the variability of the data. Chalmers attempts to distinguish between causes of the observed associations among different subsets of the components of the data, and whether these causes are due to the physical nature of the variables themselves, or to some inherent structure in the data as captured by the underlying principal components. He uses an empirical approach to generate other correlation matrices with eigenvalues identical to those observed, and then compares results from these matrices with those from the original data. The algorithm itself is derived from a geometric lemma which asserts the existence of an infinite set of orthogonal generators to certain quadratic cones in real n-space. Normalizing these generators then leads to the columns of an orthogonal matrix which transforms a given diagonal matrix of eigenvalues into the desired correlation matrix.

Bendel and Mickey (1978, Reference 12) address the same problem as Chalmers, but more systematically, and with more concern for whether the resulting correlation matrices are truly 'representative' of the entire class of correlation matrices with given spectrum, thought of as those which could arise from a given experiment. Bendel and Mickey note that parameterizing subsets of $\Gamma(N)$ by structure, e.g., equi-interclass correlation (constant off-diagonal entries) or first order autoregressive (Markov-1 data) leads to very narrow classes of correlation matrices, unsuitable for many applications. Their approach is to treat the eigenvalues as parameters, especially when they in turn are functions of one or two parameters. For example, the eigenvalues might be required to form a geometric progression. In general, if the parameterized eigenvalues are roughly constant, and therefore approximately equal to 1, the data variables are approximately independent, while a large spread in the range of the eigenvalues indicates strong interdependence between the variables.

Starting with a spectrum $\{\lambda_1, \ldots, \lambda_N\}$ and setting $D = \text{diag}\ \{\lambda_1, \ldots, \lambda_N\}$, the method of Bendel and Mickey yields a correlation matrix C of the form

$$C = U^*DU$$

where $U = VR_{N-1} \ldots R_2R_1$. Here V is a randomly chosen orthogonal (or unitary) matrix and the R's are matrices representing Givens rotations, chosen successively to make one diagonal entry at a time of the product equal to 1. The V's can be generated by various procedures; see the references[4,5] already mentioned in Section 1.2. Bendel and Mickey go on to describe the application of their method to the problem of stopping rules in the statistical procedure known as stepwise regression. They also offer some comparisons between their method and that of Chalmers.

Johnson and Welch (1980, Reference 13) also emphasize the use of simulated data to test alternative selection procedures in stepwise regression, particularly to build confidence in the use of such procedures on real data with uncertain structure. If the joint distribution of the dependent and regressor variables is Gaussian then it is standard to sample randomly from it by factoring the covariance matrix and using a string of pseudorandom N(0,1) variates (c.f. IMSL subroutines GGNSM). Thus only the structure of the distribution remains to be specified, and this, of course, is completely determined by the (mean and) covariance. If this is assumed, as they do, to be of correlation type, then it can be partitioned as

$$\begin{bmatrix} 1 & \rho^* \\ \rho & C \end{bmatrix}$$

where C is the intercorrelation matrix of the regressors, and ρ is the vector of correlation coefficients between the regressors and the dependent variable. So the emphasis is on generating such C's, and this is done by viewing C as a Gram matrix: $C = TT^*$, with the rows of T being unit vectors. Johnson and Welch suggest generating each entry of T from a symmetric beta distribution, varying the free parameter from row to row. They note that a certain control over some aspects of the matrices so defined can be maintained, such as the degree of correlation between some regressors, and the coefficient of determination for the complete regressor set.

Finally, Marsaglia and Olkin (1984, Reference 10) give a rigorous mathematical description of Methods 2 through 4 described in the preceding section. Their major result is to obtain the explicit form of the distribution of the entries of a random Gram correlation matrix $C = TT^*$, when the entries of T are generated in a particular fashion. Some of this work will be reviewed later, in the appropriate context.

2. TWO PRINCIPAL METHODS

As noted in Section 1.2, only the methods labeled as (3) and (4) are to be discussed in any detail herein. We begin this discussion now, setting the stage for the new results which will be presented later on, after a review of some salient linear algebra (see Section 3).

2.1 Random Spectrum

As we know, the spectrum of a correlation matrix $C \in \Gamma(N)$ has a spectrum $\sigma(C) = \{\lambda_1,...,\lambda_N\}$ consisting of N positive numbers of sum N. The set of all such N-tuples defines a simplex S_N, and we first want to observe that every point in S_N occurs in this fashion, that is,

$$\cup\{\sigma(C): C \in \Gamma(N)\} = S_N \quad .$$

This is a consequence of a general characterization of Hermitian matrices due to several authors. Namely, if A is a Hermitian matrix of order N with eigenvalues $\lambda_1 \geqslant ... \geqslant \lambda_N$, and diagonal entries $d_1 \geqslant ... \geqslant d_N$, then

$$d_1 + ... + d_k \leqslant \lambda_1 + ... + \lambda_k \tag{2.1}$$

for $1 \leqslant k \leqslant N - 1$, and

$$d_1 + ... + d_N = \lambda_1 + ... + \lambda_N = tr(A) \tag{2.2}$$

(Schur, 1923). Conversely, given real numbers $\{d_i,\lambda_i\}$ satisfying all these conditions, there exists a real symmetric matrix A with diagonal entries $d_1,...,d_N$, and eigenvalues $\lambda_1,...,\lambda_N$ [Horn (Reference 14), Mirsky (Reference 15)]. In our case, however, the result follows more directly from a theorem of Fillmore,[16] namely that any matrix A is unitarily equivalent to one with a constant diagonal. This in turn is an easy consequence of the convexity of the numerical range W(A), so that $tr(A)/N \in W(A)$, and an induction argument.

The upshot of the above paragraph is that given $(\lambda_1,...,\lambda_N) \in S_N$ there is a correlation matrix C with these numbers as its spectrum. How, in practice, is such a matrix to be obtained? As already noted, answers have been given by Chalmers and Bendel-Mickey; there is also the paper by Chan and Li[17] which more generally provides an algorithm for constructing a real symmetric matrix with given diagonal entries and eigenvalues satisfying the conditions (2.1) and (2.2). It appears that for present purposes the most natural method of obtaining the matrix is that proposed by Bendel and Mickey, namely,

$$C = R^*_{N-1}...R^*_2\ R^*_1\ D\ R_1\ R_2...R_{N-1} \tag{2.3}$$

where $D = diag[\lambda_1,...,\lambda_N]$, and R_k is a rotation in the plane spanned by the standard unit vectors ℓ_k and ℓ_{k+1}. The matrix R_k has the form

$$\begin{bmatrix} I_{k-1} & & & \\ & c & s & \\ & -s & c & \\ & & & I_{N-k-1} \end{bmatrix}$$

with $c^2 + s^2 = 1$. The rotation angle, arc cos(c), is chosen so that the k-th diagonal entry of C is 1

We can strengthen the preceding remark by replacing the diagonal matrix D in Equation (2.3) by $A = U^*DU$, U unitary. That is, A is an arbitrary positive definite matrix with spectrum $\{\lambda_1,...,\lambda_N\}$. Then A can still be transformed into a correlation matrix C, as before, and there are, in fact, exactly 4 choices for each R_k in Equation (2.3).

To see this, consider first the specification of R_1. $R^*_1 A R_1$ should have a diagonal entry = 1. Let A_p be a principal 2×2 submatrix of A, say

$$A_p = \begin{bmatrix} a & b \\ \bar{b} & d \end{bmatrix}$$

with a, $d > 0$. Let Q be a 2×2 orthogonal matrix, either a rotation

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

or a reflection

$$\begin{bmatrix} c & s \\ s & -c \end{bmatrix}$$

Then the upper left entry of Q^*A_pQ is $ac^2 \mp 2Re(b)cs + ds^2$, respectively. From the behavior of the Rayleigh quotient of A_p, we see that this quadratic form in (c,s) will somewhere assume the value 1 if and only if A_p has one eigenvalue $\leqslant 1$ and the other $\geqslant 1$. Now since a, d are diagonal entries of A, and tr(A) = N, A_p can be chosen so that one of a, d is $\leqslant 1$, and the other $\geqslant 1$. Its eigenvalues $\lambda_1 \geqslant \lambda_2$ then satisfy

$$\lambda_2 = \min\langle A_p x,x\rangle \leqslant \min(a,d)$$

$$\leqslant 1 \leqslant \max(a,d) \leqslant \max\langle A_p x,x\rangle \leqslant \lambda_1 \quad .$$

So the condition on A_p is satisfied, and therefore 4 choices of Q exist to yield a 1 in the upper left corner of Q^*A_pQ. R_1 is then created immediately as a direct sum of Q and an identity matrix. The whole procedure can then be repeated, since the sum of the remaining diagonal entries of $R^*_1 A R_1$ is N - 1.

The preceding remarks are a slight elaboration of some made at the end of Section 4 of Reference 10. We note that in the discussion so far of this section there are no issues of randomness. These can be introduced in two stages. First, if a point $(\lambda_1,...,\lambda) \epsilon S_N$ is given, we can form the corresponding diagonal matrix D, select an orthogonal matrix V at random from the orthogonal group O(N) with normalized Haar measure, and select a succession of orthogonal matrices, one of 4 choices at random at each step, so as to transform V^*DV into a correlation matrix C. This C may fairly be said to be a random correlation matrix with specified spectrum. Second, the spectrum may itself be chosen from some probability distribution of S_N. The resulting matrices are said to have a *random spectrum*. The special case where the distribution of S_N is uniform will be discussed at some length in Section 4. Some issues here are that this method is evidently

rather computationally expensive, and the distribution of the entries of the resulting correlation matrices is not well understood. However, we will be able to say something about the distribution of some global features of these matrices.

2.2 Random Gram Matrix

We first quickly review the concept of Gram matrix. Let $x_1,...,x_N$ be linearly independent vectors in any pre-Hilbert space. The corresponding *Gram matrix* is the $N \times N$ Hermitian matrix $G = G(x_1,...,x_N)$ with (i,j)-entry $= \langle x_i, x_j \rangle$. The determinant $g(x_1,...,x_N)$ is called the *Gramian* of the set $\{x_1,...,x_N\}$. Clearly the covariance matrix of a set of normalized second order random variables falls under this definition. In general, Gram matrices are positive semidefinite as follows from the formula

$$\langle G\alpha, \alpha \rangle = \| \sum \bar{\alpha}_i x_i \|^2 \geqslant 0$$

for any N complex numbers $\alpha_1, ..., \alpha_N$. Further, as already noted in Section 1.2, by Cholesky factorization, any real positive definite matrix is a Gram matrix; more generally, any complex positive semidefinite matrix has a positive semidefinite square root, and so is a Gram matrix.

The Gramians are symmetric functions of their arguments, and obey the inequalities

$$0 \leqslant g(x_1,...,x_N) \leqslant \|x_1\|^2 \ldots \|x_N\|^2 \tag{2.4}$$

with equality on the left if an only if $\{x_i\}$is linearly dependent, and equality on the right if and only if $\{x_i\}$ is orthogonal. To prove the right hand inequality we first reduce to the case that each x_i is a unit vector, and then

$$g(x_1,...,x_N)^{1/N} = (\Pi \lambda_i)^{1/N}$$

$$\leqslant \frac{1}{N} \sum \lambda_i = \frac{1}{N} \operatorname{tr}(G) = 1$$

where $\{\lambda_i\} = \sigma(G)$.

One senses from this that the Gramian and other spectral features of the Gram matrix bear some relation to the relative orientation of the vectors $\{x_i\}$. Along this line we recall that if the vectors x_i belong to R^N, then

$$\operatorname{vol}\left(\left\{ \sum_1^N \lambda_i x_i : 0 \leqslant \lambda_i \leqslant \epsilon_i \right\} \right) =$$

$$g(x_1,...,x_N)^{1/2} \prod_1^N \epsilon_i \tag{2.5}$$

so that in particular $g(x_1,...,x_N)$ is the square of the volume of the parallelepiped spanned by the set $\{x_i\}$. If the x_i belong to some other space, Equation (2.5) serves to define this volume.

In addition to the simple Hadamard inequality in Equation (2.4), we have further

$$g(x_1,...,x_m,y_1,...,y_N) \leq g(x_1,...,x_m)\ g(y_1,...,y_N)$$

and in fact the ratio of the left side to the right side is known to be $\sin^2 \alpha_1 ... \sin^2 \alpha_M$, where $\alpha_1,...,\alpha_M$, $M \leq N$, are the angles between span$\{x_i\}$ and span$\{y_j\}$.

Gram matrices occur naturally in all manner of least squares problems, such as Gram-Schmidt orthogonalization, linear regression, and pseudoinversion. Indeed, the basic problem of computing the orthogonal projection onto span $\{x_i\}$ requires the solution of a linear system with Gram coefficient matrix. It is familiar that as the basis vectors x_i deviate more from orthogonality, such problems become more ill-conditioned, and associated statistical procedures are said to suffer from 'collinearity.' For example, if $x_i(t)$ is the monomial t^i, and the inner product comes from some Lebesque-Stieltjes measure with compact support, then the corresponding Gram matrices, indexed by N, have a condition number that grows at least as fast as 4^N; the classical Hilbert matrices are special cases;[18] the main result of this reference will be discussed later in Section 3. For a recent review of the collinearity problem with suggestions for its measurement by more subtle indicators than simply condition number, see Reference 19.

As a reference point for later use, we record here a well-known distance formula involving Gramians. Let $M = \text{span}\ \{x_1,...,x_N\}$, and let x be another point in the space containing M. Then

$$\text{dist}(x,M)^2 = \frac{g(x_1,...,x_N,x)}{g(x_1,...,x_N)} \quad . \tag{2.6}$$

Recall that random Gram matrices were defined by Method 4 in Section 1.2. In present notation the x_i are taken to be random vectors uniformly distributed over the sphere S^{N-1} in R^N. We now report some results from a small simulation, intended to compare such matrices with those of random spectrum (Method 3). We give here only the cases N = 5 and 10, as they appear typical of all cases considered. In each case, the summary statistics are based on 1000 trials. In the left column of Table 1, 'c.n.' means condition number, 'F norm' means Frobenius norm, and 'norm' means spectral norm. Also, 'trimmed' means that the largest 1 percent and smallest 1 percent of the samples have been deleted.

Probably the most striking contrast to be made on the basis of this numerical experiment is the excessively high condition numbers of the random Gram matrices relative to those of the matrices with random spectrum. This aspect of the data persists even after trimming, and after passing to medians. It strongly suggests that random Gram matrices do not have a random spectrum. It also raises interesting questions about the relative orientation of a batch of 2 or more vectors drawn independently from the uniform distribution on the (N - 1)-sphere. Some of these will be considered in Section 5 below.

TABLE 1

Statistics for Random N × N Matrices

	N = 5	10	N = 5	10
Mean c.n.	111.	553.	1.37E6	7.85E6
Standard Deviation	754.	1.18E4	4.29E8	2.19E8
Median c.n.	15.8	40.5	114.	809.
Interquartile Range	30.7	80.9	446.	3533.
Trimmed Mean	39.2	95.4	1.57E3	1.57E4
Standard Deviation	70.3	157.	5.48E3	6.65E4
Mean F. Norm	2.88	4.24	2.99	4.36
Mean Norm	2.31	2.92	2.43	2.94
Mean Min. E-Value	.191	.100	.042	.009
Standard Deviation	.161	.093	.054	.012
	Random Spectrum		**Random Gram**	

3. ASPECTS OF NUMERICAL LINEAR ALGEBRA

This section contains a brief review of some quantitative aspects of linear algebra that are pertinent to the material that follows. For general background information on matrix theory we may refer to two recent volumes: Horn and Johnson[20] or Lancaster and Tismenetsky.[21] More specialized treatments of numerical linear algebra are given by Stewart[22] and Golub-van Loan.[23]

3.1 Bounds on Norms and Eigenvalues

Given an $N \times N$ matrix A we shall have occasion to use its operator or spectral norm

$$\|A\| = \max\{\|Ax\| / \|x\|: x \neq \theta\}$$

and its Frobenius norm

$$\|A\|_F = (\sum |a_{ij}|^2)^{1/2} \quad .$$

In terms of the positive part $P = (AA^*)^{1/2}$ of A, we have

$$0 \leqslant \|A\| = r_\sigma(P) \leqslant \|A\|_F = \mathrm{tr}(P) \tag{3.1}$$

where $r_\sigma(\cdot)$ means spectral radius. Bringing in the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_N$ of A, and the singular values $s_1 \geqslant \ldots \geqslant s_N$ (these are the eigenvalues of P), we have $\|A\| = s_1$, and

$$\sum_1^N |\lambda_i|^2 \leqslant \|A\|_F^2 = \sum_1^N s_i^2 \tag{3.2}$$

with equality if and only if A is normal (theorem of Schur and Mirsky).

For general matrices A, the singular values have many fascinating properties and applications, such as min-max characterizations, smooth dependence on A (which leads into perturbation theory), and geometric interpretations as distances from A to spaces of matrices of lower rank. This latter property, on the one hand, leads into regularization techniques for least squares signal processing and, on the other, permits generalization to compact operators on infinite dimensional spaces (s-number theory).

Let us now specialize to the case of primary interest here, namely, that where $A = C$ is a correlation matrix. Then we have

$$\begin{aligned} 1 \leqslant \|C\| = r_\sigma(C) &= \min\{\|C\|\} \\ &\leqslant \max\{\|\mathrm{row}_1\|_1, \ldots, \|\mathrm{row}_N\|_1\} \leqslant N \end{aligned} \tag{3.3}$$

where $|||\cdot|||$ refers to a general matrix norm induced by some vector norm, and $\|\cdot\|_1$ is the ℓ^1-vector norm. The expression "$\max\{\ldots\}$" above is just the matrix norm induced by the ℓ^∞-vector norm. Its advantage, as with the bound $\|C\|_F$, is that it is immediately computable from the entries of C. Either of the extremes 1, N can be reached by some $C \in \Gamma(N)$. The second equality is true in much greater generality; in fact, it is true for any operator on a Banach space.[24]

If A is positive semidefinite, then

$$|a_{ij}| \leqslant \sqrt{a_{ii}a_{jj}} \leqslant \frac{1}{2}(a_{ii} + a_{jj})$$

showing in particular that all off-diagonal entries of a correlation matrix have modulus $\leqslant 1$. Of course, such a matrix need not be diagonally dominant.

An improvement on the bound $\|C\| \leqslant \|C\|_F$ has been noted by Leclerc,[25] specifically for correlation matrices. Namely,

$$\|C\| \leqslant 1 + \left(\frac{N-1}{N} \sum_0^2\right)^{1/2} \leqslant \|C\|_F \tag{3.4}$$

where $\sum_0^2$ = sum of squares of off-diagonal entries of C. The right-hand inequality here is strict unless all off-diagonal entries have modulus = 1. This bound on $\|C\|$ can be either larger or smaller than the 'max' bound of Equation (3.3).

At this point we have given some upper bounds for $\|C\|$, and hence for all the (positive) eigenvalues of C. Upper bounds for $\|C^{-1}\|$ are equivalent to lower bounds on the eigenvalues of C, using $\|C^{-1}\| = r_\sigma(C^{-1})$; note that (C^{-1}), while still positive definite, is no longer a correlation matrix in general. This kind of bound is not of particular interest to us here, but lower bounds on $\|C^{-1}\|$ are important, in connection with condition number estimates, and will be discussed later on. Here we will just recall an inequality of Kato,[26] which gives abound on $\|A^{-1}\|$, for any nonsingular A:

$$\|A^{-1}\| \leqslant \|A\|^{N-1}/|\det(A)| \quad .$$

There are innumerable inequalities pertaining to the eigenvalues of positive definite matrices (and operators). We mention just two, for products and sums of eigenvalues, due to K. Fan. With A positive definite and its spectrum ordered as $\lambda_1 \geqslant \ldots \geqslant \lambda_N$:

$$\lambda_N \lambda_{N-1} \ldots \lambda_k = \min \prod_{i=k}^{N} \langle Ae_i, e_i \rangle$$

$$\lambda_1 + \ldots + \lambda_k = \max \sum_{i=1}^{k} \langle Ae_i, e_i \rangle$$

$k = 1, \ldots, N$, where the min (respectively max) is taken over all sets of $n - k + 1$ (respectively k) orthonormal vectors (see Reference 27, Chapter 2, for these and many others).

Finally, we mention the concept of *spread* of a matrix A. This is the quantity

$$S(A) = \text{diam } \sigma(A) \equiv \max|\lambda_i - \lambda_j| \quad . \tag{3.5}$$

When A = C, a correlation matrix, the following bounds on S(C) can be derived;

$$2 \max_{i \neq j} |c_{ij}| \leqslant S(C) \leqslant (2(\|C\|_F^2 - N))^{1/2} \quad .$$

Since $\|C\|_F^2 = \Sigma \lambda_i^2$, the last inequality offers a lower bound on this quantity. But, in fact, a stronger 2-sided inequality can be established, namely

$$\frac{1}{2} S(C)^2 + N \leqslant \|C\|_F^2 \leqslant \frac{5}{4} N S(C)^2$$

by working with eigenvalues.

3.2 Condition Number Estimates

The condition number $\kappa(A)$ of an arbitrary matrix A is defined by

$$\kappa(A) = \|A\| \ \|A^+\| = \|A\| \ \|A^{-1}\| \tag{3.6}$$

where the '+' means pseudoinverse, and the second equality is naturally only applicable if A is nonsingular. Note that this is the spectral condition number; other matrix norms might be used in Equation (3.6). In terms of the singular values of A we have

$$1 \leqslant \kappa(A) = s_1 / s_N \tag{3.7}$$

with equality if and only if A is a nonzero multiple of a unitary matrix. [Naturally, Equation (3.7) is restricted to nonsingular A.] Many kinds of singular matrices A can have $\kappa(A) = 1$; for instance, orthogonal projections and, more generally, partial isometries.

Condition numbers are widely used as measures of sensitivity of the solution of linear systems to inaccuracies in the data. Similarly, the condition number of the matrix of eigenvectors of a diagonalizable matrix measures the closeness of an approximate eigenvalue to the true spectrum. Roughly speaking, the percentage change in the (least squares) solution x of the system $Ax = b$ is bounded by the percentage variation in the data b times $\kappa(A)$, and this bound cannot be lowered. Thus $\kappa(A)$ is a measure of the inherent resistance of a particular system to accurate solution, and which does not depend on the particular numerical method employed. The larger the condition number, the more 'ill conditioned' a particular system is, and the less we can infer a small error from a small residual.

We might also remark that $\kappa(A)$ can be characterized geometrically by the least angle ψ resulting as A is applied to all possible pairs of orthonormal vectors; precisely

$$\kappa(A) = \cot(\psi / 2) \quad .$$

It is instinctive to want to measure ill conditioning by some function of the eigenvalues, but this is only fruitful for normal matrices. For example, there is the $N \times N$ 'Kahan matrix'

$$\begin{bmatrix} 1 & -1 & . & . & . & -1 \\ 0 & 1 & . & & & \\ . & & & . & & \\ . & & & & . & \\ . & & & & & -1 \\ 0 & . & . & . & & 1 \end{bmatrix}$$

which clearly has all eigenvalues = 1, yet a condition number $> \sqrt{N}2^{(N-2)}$. However, when A is positive definite with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$, then

$$\kappa(A) = \lambda_1 / \lambda_N \tag{3.8}$$

and we have the inequality of Kato:

$$\kappa(A) \leq \frac{4}{\det(A)} \left(\frac{\mathrm{tr}(A)}{N} \right)^N .$$

Thus, for correlation matrices C, $\kappa(C)\det(C)$ is a bounded function. Note that Equation (3.8) defines a different notion of 'spread' of the spectrum, in contrast with the quantity S(A) of Equation (3.5). Also note that Equations (3.7) and (3.8) together imply that

$$\kappa(A^*A) = \kappa(AA^*) = \kappa(A)^2$$

showing how the familiar 'normal equations' of many least squares procedures can become very ill-conditioned (and, eventually, motivating the use of factorization methods which deal only with A, as an alternative).

In 1955 J. Riley[28] used the fact that

$$\kappa(A + \lambda I) \leq \kappa(A) \tag{3.9}$$

for any positive definite A to suggest an iterative improvement procedure for solving an ill-conditioned linear system $Ax = b$. This was a forerunner of the ridge regression and regularization methods of statistics and signal processing, which trade off some bias for lowered mean square error. The inequality (3.9) was greatly extended by Marshall and Olkin[29] who proved that

$$\kappa(A + B) \leq \kappa(A)$$

whenever A and B are positive definite with $\kappa(B) \leq \kappa(A)$.

We now turn to the matter of *lower* bounds for condition numbers. These will be of greatest interest for the case of Gram matrices, but we consider first, briefly, the general case. (We note, too, the considerable interest in recent years in numerical estimates - not bounds - for condition numbers, by estimating some norm of the inverse matrix.[30,31,32]

First, if A is any nonsingular matrix, with eigenvalues ordered by modulus: $|\lambda_1| \geq \ldots \geq |\lambda_N|$, then

$$|\lambda_1 / \lambda_N| \leq \kappa(A) \quad . \tag{3.10}$$

This follows from the relations

$$\|A^{-1}\|^{-1} = \inf\{\|Ax\|: \|x\| = 1\}$$
$$\leqslant \|Ae\| = |\lambda|$$

where e is any unit eigenvector associated with $\lambda \epsilon \sigma(A)$. Of course, as the earlier example of the Kahan matrix illustrates, the left side of Equation (3.10) may severely underestimate the true condition number $\kappa(A)$, when A is not normal.

Now assume that A is positive definite. A variant of the well-known Kantorovich inequality[33] tells us that

$$\|x\|^2 \leqslant \langle Ax,x\rangle \langle A^{-1}x,x\rangle \leqslant \frac{(m_1 + m_2)^2}{4m_1 m_2} \|x\|^2 \tag{3.11}$$

provided that

$$m_1 I \leqslant A \leqslant m_2 I$$

for $0 < m_1 \leqslant m_2$. Taking m_1 (respectively m_2) to be the least (respectively greatest) eigenvalue of A, and x any unit vector, we obtain

$$4\langle Ax,x\rangle \langle A^{-1}x,x\rangle \leqslant \kappa + \frac{1}{\kappa} + 2 \leqslant \kappa + 3$$

yielding a lower bound for $\kappa = \kappa(A)$ for each x. Of course, an estimate involving A^{-1} is not of great practical value.

Another kind of inequality comes from the theory of Schur (or Hadamard) products of matrices. We won't review this concept in any detail here; see Reference 34 for a nice survey. This product, for conformable matrices A, B, is defined by

$$[A \cdot B]_{i,j} = a_{ij} \cdot b_{ij} \quad .$$

This multiplication, unlike the usual one, is commutative. The original result of Schur is that if A, B are positive semidefinite, then so is $A \cdot B$. An inequality of Fiedler (1961) for positive definite A reads

$$A \cdot A^{-1} \geqslant I \quad . \tag{3.12}$$

Note that, as a consequence of either this or the left side of the Kantorovich inequality, when C is a correlation matrix,

$$[C^{-1}]_{i,i} \geqslant 1, \qquad i = 1, \ldots, N \quad .$$

In 1982 M. Marcus proved the matrix norm inequality

$$\|A \cdot B\| \leqslant \|A\| \, \|B\|$$

for the Schur product. Taking $B = A^{-1}$ yields a lower bound for the condition number:

$$\|A \cdot A^{-1}\| \leqslant \kappa(A)$$

In preparation for part of our discussion in Section 5, we want to consider specifically the case where A is a Gram matrix:

$$A = G = G(x_1, \ldots, X_N)$$

in the notation of Section 2.2, with each x_i a unit vector in some inner product space X. As already remarked in Section 2.2 it has been empirically noted that many common Gram matrices tend to be ill-conditioned, and an inequality derived in Reference 18 can be used to quantify these observations by providing a lower bound for $\kappa(G)$ in terms of the relative orientations of the vectors $\{x_i\}$. By virtue of our own numerical experiments reported earlier, ill conditioning is a prominent feature in random Gram matrices also. We will now discuss an improved version of this inequality, and its sharpness. These results are purely deterministic; statistical implications are deferred to Section 5.

A fixed Gram matrix $G = G(x_1, \ldots, x_N)$, $\|x_i\| = 1$, $i = 1, \ldots, N$ will be used. G is a correlation matrix, and $\|G\| = r_\sigma(G)$, so the problem is to find a lower bound on $\|G^{-1}\|$ in terms of the vectors $\{x_i\}$. Let M (respectively M_i) be the subspace of X spanned by $\{x_j\}_1^N$ (respectively $\{x_j : j \neq i\}_1^N$). Let $\{v_j\}$ be the basis for M that is dual to $\{x_j\}$. Also, for an arbitrary real or complex unit vector e (X is real or complex), let $b = G^{-1}e$. Then

$$\|G^{-1}\| \geq \langle G^{-1} e, e\rangle = \langle b, Gb\rangle$$
$$= \|\sum \bar{b}_i x_j\|^2 \equiv \|v\|^2 \quad .$$

Now, with v as just defined, it is easily checked that

$$\bar{b} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \langle x_i, v\rangle \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

so that if $v = v_i$, one of the dual basis vectors in M, then $Gb = e_i$, the standard unit basis vector.

We also observe that, since M_i is of codimension 1 in M, the duality formula for distance,

$$\text{dist}(x, M_i) = \max\{|\psi(x)| : \psi \epsilon S(M_i^\perp)\} \tag{3.13}$$

for $x \epsilon M$, implies that

$$d_i \equiv \text{dist}(x_i, M_i) = \langle x_i, v_i / \|v_i\|\rangle$$
$$= 1/\|v_i\| \quad . \tag{3.14}$$

Putting all this together, we conclude that

$$\|G^{-1}\| \geq \langle G^{-1} e_i, e_i\rangle = \|v_i\|^2$$
$$= 1/d_i^2 = \frac{g_i(x_i, \ldots, x_N)}{g(x_1, \ldots, x_N)}$$

where the last equality follows from the Gramian distance formula of Equation (2.6), and 'g_i' means the Gramian with x_i deleted.

The ensuing inequality

$$\|G^{-1}\| \geq \max d_i^{-2} = \frac{1}{\min d_i^2} \tag{3.15}$$

is due to J. Taylor (Reference 18, p. 46). The major difference between Taylor's approach and the one we are using is that the duality formula Equation (3.13) strengthens the inequality by avoiding reliance on the Schwarz inequality. Thus the sole source of inequality in Equation (3.15) is the inequality appearing in Equation (3.14). This inequality is only a measure of the behavior of the Rayleigh quotient for G^{-1}, and does not explicitly involve the Gram structure of G. Hence the following theorem and proof gives a measure of the tightness of Taylor's inequality Equation (3.15).

Theorem

$$\sup_{A \in \Gamma(N)} \frac{\|A^{-1}\|}{\max \langle A^{-1}e_i, e_i \rangle} = N.$$

Proof For notational ease, we will replace A^{-1} by A, and then $\max \{\langle Ae_i, e_i \rangle : i = 1, \ldots, N\}$ by $\mu(A)$. We first note that if A is any $N \times N$ positive definite matrix,

$$1 \leq \|A\| / \mu(A) \leq N$$

and that these bounds are sharp (within this larger class of matrices). The left inequality is trivial, and is achieved for diagonal matrices. The right inequality follows from

$$\|A\| = r_\sigma(A) = \lambda_1 \leq \mathrm{tr}(A) \leq N\mu(A) \quad .$$

To verify its sharpness, let $\epsilon > 0$, and $D = \mathrm{diag}(1, \epsilon, \ldots, \epsilon)$, and apply the theory in Section 2.1 to obtain A, unitarily equivalent to D, with constant diagonal. Then

$$1 = \|A\| \leq \mathrm{tr}(A) = N\mu(A) = 1 + (N - 1)\epsilon$$

now let $\epsilon \downarrow 0$. So the point of the theorem is that if the A's are restricted to the class $\{A: A^{-1} \in \Gamma(N)\}$, the upper bound on $\|A\| / \mu(A)$ does not decrease.

To complete the proof, consider a special family of A's, namely, $\{A: A = aI_N + B, b_{ij} = (1 - \delta_{ij})b\}$, where $a > b > 0$. We have

$$\begin{aligned} \|A\| / \mu(A) &= (a + (N - 1)b)/a \\ &= 1 + (N - 1)\, b/a \end{aligned} \tag{3.16}$$

and it will be shown that a and b can be chosen so that $A^{-1} \in \Gamma(N)$ and

$$\lim_{a \uparrow \infty} \frac{b}{a} = 1. \tag{3.17}$$

Let $\Delta = \det(A)$. We have

$$\Delta = (a - b)^{N-1}(a + (N - 1)b)$$

and since the diagonal entries of A^{-1} are

$$\langle A^{-1}e_i, e_i \rangle = i^{th}\text{-cofactor} / \Delta$$

it follows that

$$\langle A^{-1}e_i, e_i \rangle = \frac{(a + (N - 2)b)(a - b)^{N-2}}{(a + (N - 1)b)\ (a - b)^{N-1}} \quad .$$

So, if A^{-1} is to be a correlation matrix, a and b must satisfy the equation

$$a + (N - 2)b = (a - b)\ (a + (N - 1)b).$$

If we treat this as a (quadratic) equation for b and solve it, we obtain

$$b = \frac{(a - 1)\ (N - 2) + ((N - 2)^2(a - 1)^2 + 4(N - 1)\ (a^2 - a))^{1/2}}{2(N - 1)} \quad .$$

After dividing both sides by a, and manipulating, we have

$$\frac{b}{a} = 1 - \frac{1}{a}\ \frac{N - 2}{2N - 2} + \frac{\sqrt{(N^2 + e)} - N}{2N - 2}$$

where

$$e = \frac{(N - 2)^2}{a}\left(\frac{1}{a} - 2\right) - \frac{4}{a}(N - 1) < 0 \quad .$$

This shows that $b < a$ and that the limit in Equation (3.17) is 1, as required. QED

At this point we might justify an assertion made just after Equation (3.3), namely, that

$$\boxed{\sup_{A \in \Gamma(N)} \|A\| = N \quad .}$$

We know from that equation that this supremum is at most N. That it is not less than N follows from consideration of the same family of matrices just used, and the value of the norms of such matrices given in Equation (3.16): just take $a = 1$ and let $b \uparrow 1$.

To sum up: for a Gram matrix $G = G(x_1, \ldots, x_N)$, we have the lower bound on $\kappa(G)$, due to Taylor:

$$\kappa(G) \geq 1 / \min_i d_i^2$$

an equivalent form

$$\kappa(G) \geq \max_i \frac{g_i(x_1, \ldots, x_N)}{g(x_1, \ldots, x_N)}$$

and an upper bound on the tightness of this lower bound:

$$1 \leqslant \frac{\kappa(G)}{\text{lower bound}} \leqslant N^2 \quad . \tag{3.18}$$

It is possible that this upper bound could be decreased, but we haven't investigated this point. Some evidence is given below in Section 5.4.

And so, our review and development of numerical linear algebra now complete, we return to discuss our primary topic where the matrices have a random nature.

4. CORRELATION MATRICES WITH RANDOM SPECTRUM

Some background for this section was given in Section 2.1. There, two essential steps in defining correlation matrices with random spectrum were recognized:

- pick a point $\underline{\lambda} = (\lambda_1,...,\lambda_N)$ 'at random' from the simplex S_N, and form the diagonal matrix $D = \text{diag}(\lambda_1,...,\lambda_N)$
- construct a matrix $C = R^*V^*DVR$, where V is drawn at random from the orthogonal group O(N), and R is a product of randomly selected rotation/reflection matrices, chosen to successively put 1's on the diagonal of C.

Naturally the second step leaves $\sigma(C) = \underline{\lambda}$, and hence leaves all spectral functions of C unchanged. Among such functions are the spectral and Frobenius norms of C, and its condition number. (In general, any unitarily invariant norm of a Hermitian matrix is a spectral function of that matrix.) Consequently, once a probability distribution is selected on S_N, so as to define the 'at random' condition of the first step, the spectral functions of any correlation matrix defined by the second step may be directly studied. Note that this approach does not apply to other numerical attributes of C of possible interest, such as its individual entries; their distribution naturally depends in part on those of the V and R matrices.

4.1 Method of Generation

From the preceding discussion we see that a probability distribution μ must be specified on the simplex S_N. Then a sample $\underline{\lambda}$ drawn from μ will be a random spectrum. Having no reason to weight any region of S_N more than another, we will take μ to be the uniform distribution on S_N, and denote by

$$\underline{\lambda} \sim U(S_N)$$

a point $\underline{\lambda}$ so chosen. Two tasks remain:

- specify μ analytically
- specify μ operationally

This latter task simply means to prescribe a method for a computer to make these draws in terms of an assumed capability to generate pseudorandom numbers $\sim U[0,1]$.

The analytical specification of μ depends on (a special case of) the general theory of order statistics and spacings. Here we only need the case of independent samples from the uniform distribution. Thus let $u_{(1)} \leqslant ..... \leqslant u_{(N-1)}$ be the order statistics of a random sample from U([0,1]). Define $u_{(0)} = 0$, and $u_{(N)} = 1$. As shown by Wilks,[35] the joint distribution of these order statistics is uniform over the simplex $\{0 \leqslant x_1 \leqslant ... \leqslant x_{N-1} \leqslant 1\}$ in R^{N-1}. Then the *spacings* of the sample are defined by

$$s_i = u_{(i)} - u_{(i-1)}, \qquad 1 \leqslant i \leqslant N \quad .$$

Observe that for each sample, the spacings are positive numbers that sum to 1. It was also shown by Wilks that the spacings vector $(s_1,...,s_{N-1})$ is uniformly distributed over the simplex

$$\{0 \leqslant x_i, \sum_{1}^{N-1} x_i \leqslant 1\}$$

in R^{N-1}. Now, for fixed N, the mapping

$$T(x_1,...,x_{N-1}) = (x_1,...,x_{N-1}, 1 - x_1 - ... - x_{N-1})$$

carries this simplex bijectively onto the simplex $\{0 \leqslant x_i, \sum x_i = 1\}$ in R^N, and carries over the uniform density (by an elementary change of variables).

The upshot of the preceding paragraph is that, for fixed N, the uniform density μ on the simplex S_N can be specified analytically as the distribution of the random vector

$$NT(s_1,...,s_{N-1}). \tag{4.1}$$

And, of course, it follows that μ can be specified operationally in terms of pseudorandom number generator, and a sorting routine.

There is by now a fairly extensive literature of spacings (sometimes known as 'gaps', 'coverages', or 'random division of an interval'). This topic can be traced back to the turn of the century to the work of W. Whitworth on the distribution of the largest spacing. His result was utilized by Fisher (1929) to give a significance test for the largest amplitude in a numerical harmonic analysis of a time series. (In fact, Fisher's test turns out to be the uniformly most powerful symmetric invariant decision procedure against simple periodicities. More recent work is concerned with compound periodicities, and hence with the distribution of other functions of the spacings besides the maximum. Interested readers should consult the papers of A. Siegel, for example, Reference 36. We will not detail any of his work here, we merely wanted to draw attention to the unexpected link between the spacings concept and time series analysis.)

In the later 1930's and then the 1940's several authors including P. Levy, M. Greenwood, and P. Moran worked on distributions of functions of spacings. Most of this work originated as specific problems in applied statistics. The best review of all this is that of R. Pyke,[37] although at a fairly high technical level. Other useful works include References 38 and 39. These references point out, among many other things, alternative analytical specifications of spacings. For instance, if $y_1,...,y_N$ are independently exponentially distributed with arbitrary mean, and $z = y_1 + ... + y_N$, then the random vector

$$z^{-1}(y_1,...,y_N)$$

is distributed as the spacings vector $T(s_1, ..., s_{N-1})$. Hence random spectra can also be generated by use of exponential variates. From this it follows that spacings can also be simulated from the

(normalized) interarrival times of a Poisson process $\{N(t): t \geq 0\}$ with $N(0) = 0$. Namely, if T_k is the elapsed time between the (k-1)-st and the k-th event, and $t > 0$ is fixed, then the conditional distribution, given $N(t) = n$, of

$$t^{-1}(T_1,...,T_{n-1}, t - T_{N(t)})$$

is the same as the spacings vector. This is a classic construction of spacings with important modern applications to the limiting behavior of empirical processes.[37]

The distribution of spacings and some functions thereof is also briefly discussed in Kendall and Moran.[40] Naturally, geometric aspects are stressed. For instance, the joint distribution of the spacings is, with proper scale factor, exactly that of the lengths of the N perpendiculars from a random point inside the simplex S_N to the N sides. The authors go on to discuss some situations where probabilities can be computed from simplicial geometry.

4.2 Distribution of Eigenvalues

Pursuant to the foregoing discussion we take as a random spectrum $\underline{\lambda} \in S_N$, N times the random vector of spacings defined by a random sample of size N - 1 from the standard uniform distribution. We denote this vector as $\underline{\lambda} = (\lambda_1,...,\lambda_N)$. In this short section we discuss the distribution of the λ_i, while in the next two sections we consider that of certain functions of the λ_i related to correlation matrices C with $\sigma(C) = \underline{\lambda}$.

We first note that the λ_i are exchangable random variables since, because of the uniform distribution of $\underline{\lambda}$ on S_N, that distribution is unchanged under permutation of its components. It follows that the λ_i are identically distributed and, using the formula for the least order statistic, the distribution function F_N is given by

$$F_N(t) = 1 - (1 - \frac{t}{N})^{N-1} . \tag{4.2}$$

Thus, for large N, λ_i is approximately exponentially distributed with mean 1. From Equation (4.2) we can conclude that

$$E(\lambda_i) = 1, \ var(\lambda_i) = \frac{N - 1}{N + 1} ,$$

for each i.

Expressions for the joint distribution of the λ_i have been given by Steutel.[39] Namely,

$$\Pr(\lambda_1 > \alpha_1,\ldots,\lambda_N > \alpha_N) = \begin{cases} (1 - \frac{1}{N}\sum \alpha_i)^{N-1}, & \sum \alpha < 1 \\ 0 & \text{, if not} \end{cases}$$

and

$$\Pr(\lambda_1 \leqslant \alpha_1,\ldots,\lambda_N \leqslant \alpha_N) = 1 - \sum_{j=1}^{N} (1 - \frac{\alpha_j}{N})^{N-1}$$

$$+ \sum_{j,k=1}^{N} (1 - \frac{\alpha_j + \alpha_k}{N})^{N-1} - + \ldots .$$

These formulae are derived by Laplace transform techniques and the relation, already alluded to in Section 4.1, between the spacings distribution and that of certain exponential variates.

In similar fashion one could go on to describe the joint distribution of pairs (λ_i,λ_j), the associated covariance, etc. Here we will just note that

$$\mathrm{corr}(\lambda_i,\lambda_j) = \frac{-1}{N-1} \quad .$$

But actually all such formulae of likely interest follow directly from the *multiple moments formula*

$$E(\lambda_1^{p_1} \ldots \lambda_N^{p_N}) = N^p \Gamma(N) \frac{\Gamma(p_1 + 1) \ldots \Gamma(p_N + 1)}{\Gamma(p + N)} \tag{4.3}$$

where $p = p_1 + \ldots + p_N$. This expression can either be derived by the Laplace transform method of Steutel,[39] or, somewhat more directly and geometrically, as in Kendall and Moran (Reference 40, page 34).

4.3 Distribution of Norms

We continue with the assumption that we are dealing with a correlation matrix C whose spectrum $\underline{\lambda}$ has been chosen randomly according to $U(S_N)$. The issue now is the distribution of the norms $||C||$ and $||C||_F$, as defined in Section 3.1.

Let us begin with $||C||_{F2} = \sum \lambda_i^2$ which, for both typographical and historical reasons, we will denote by GM(N). In the statistical literature this quantity is known as the Greenwood-Moran statistic, after the authors of References 41 and 42. It was originally proposed as a test

for uniformity in response to a problem in epidemiology (time intervals between outbreaks of an infectious disease). Moran[42] derived a general formula for the moments of GM(N); it was re-derived by Steutel.[39] For us, it is enough to use the moments formula (4.3) to obtain

$$E(GM(N)) = \frac{2N^2}{N+1}$$

$$E(GM(N))^2 = \frac{4N^4\,(N+5)}{(N+1)\,(N+2)\,(N+3)}$$

and hence

$$\mathrm{var}(GM(N)) = \frac{4N^4\,(N-1)}{(N+1)^2\,(N+2)\,(N+3)}$$

$$= O(N) \quad .$$

If now for fixed N we take a large sample of size n of random correlation matrices C with spectra $\underline{\lambda} \sim U(S_N)$, we would expect $||C||_F$ to be approximately normally distributed with mean

$$\sqrt{\left(\frac{2N^2}{N+1}\right)}$$

and variance about $1/2n$. This is a consequence of general theory concerning asymptotic distribution of continuous functions of sample means (Reference 35, page 259). In the particular cases of N = 5 and 10, we expect, in a large sample, to have ave($||C||_F$) about equal to 2.89 and 4.26, respectively. The reader may look back to the first two columns of Table 1 for the actual result of a sample of size n = 1000.

A second point to be made about GM(N) is that it is (slowly!) asymptotically normal, a result due originally to Moran,[42] and reproved by a more general method by Darling,[43] see also References 37 and 39. As usual, Pyke has the most complete but also most opaque discussion of this topic. Once this asymptotic normality is established, the corresponding property of $\sqrt{GM(N)} = ||C||_F$ can be worked out by general theory concerning smooth functions of asymptotically normal variates. In fact, since $GM(N)/2N$ has mean $N/(N+1) \approx 1$, and variance $\sigma_N^2 = O(1/N)$, its asymptotic normality implies that $\sqrt{GM(N)/2N}$ is asymptotically normal with mean 1, and variance $\sigma_N^2/4$. Hence $\sqrt{GM(N)}$ is asymptotically normal with mean $\sqrt{2N}$, and variance $(1/2)N\sigma_N^2 \approx 1/2$.

Next we consider the spectral norm $||C||$, for an $N \times N$ correlation matrix C with random spectrum as usual. Since $||C|| = \lambda_{max}$, the maximum eigenvalue of C, the distribution of $||C||$ is that of N times the maximum spacing determined by a random sample of N - 1 points from the

standard uniform distribution. We let V_N denote this maximum spacing, so that $\lambda_{max} = ||C|| \approx NV_N$.

As already noted in Section 4.1, the distribution of V_N goes back to Whitworth (1897), and has a history of interesting statistical applications. A convenient source for this distribution is Reference 43, wherein one can also find a derivation of the asymptotic behavior, due originally to Levy (1939). We find that

$$\Pr(NV_N < x) = \sum_{k=0}^{N} (-1)^k \binom{N}{k} \left(1 - \frac{kx}{N}\right)_+^{N-1}$$

where $(t)_+ = \max(t,0)$. From this one could derive the mean and higher moments, as needed. As a somewhat neater alternative, we can appeal to some well-known relations between the distribution of the spacings and certain exponential variates, as briefly reviewed in Section 4.1. Now making use of the fact that the sum of exponential variates y_i is gamma-distributed, and the known distribution of the order statistics from the exponential distribution, we can obtain

$$NV_N \sim N \max\{y_i\}/z.$$

Also, a formula was given by Devroye.[44]

$$v_N \sim \left(\sum_{1}^{N} y_i/i\right)/z$$

In both cases $z = y_1 + \ldots + y_N$. From all this we can deduce that

$$E(\lambda_{max}) = E(NV_N) = 1 + \frac{1}{2} + \ldots + \frac{1}{N}$$

$$\approx \log N + \gamma$$

where $\gamma = .577 \ldots$ is Euler's constant.

Finally, the Levy-Darling asymptotic formula for the maximum spacing, scaled to apply to the maximum eigenvalue of the matric C is

$$\Pr(NV_N < \log N + x) \to \exp(-e^{-x})$$

as $N \to \infty$. From this, it follows that

$$\text{var}(NV_N) \to \pi^2/6$$

as $N \to \infty$.

Thus we have obtained the exact means of the norm functions $||C||_F^2$ and $||C||$, and the asymptotic behavior of these, along with $||C||_F$, as $N \to \infty$. In particular, we have observed that the Frobenius norm tends to normality, while the spectral norm tends to obey an extreme value distribution.

4.4 Condition Number Expectation

We continue to study an $N \times N$ correlation matrix C with random spectrum. Now our focus is on the distribution of the condition number $\kappa(C)$, as defined by Equation (3.6). As we know from Equation (3.8), $\kappa(C) = \lambda_{max}/\lambda_{min}$, the ratio of the largest to the smallest eigenvalue of C. We have just described the distribution of $\lambda_{max} = ||C||$. In fact the *joint* distribution of (λ_{max}, λ_{min}) can be inferred from the work of Darling,[43] in the following form:

$$\Pr(\lambda_{min} > x,\ \lambda_{max} < y)$$

$$= \sum_{j=0}^{N} \binom{N}{j} (-)^j \left(1 - x\left(\frac{N-j}{N}\right) - y\,\frac{j}{N}\right)_+^{N-1} .$$

From this, by letting $y \uparrow N$, we obtain the distribution for the least eigenvalue:

$$\Pr(\lambda_{min} > x) = (1 - x)^{N-1} \qquad 0 < x < 1 \tag{4.4}$$

This formula yields the moments of λ_{min} as

$$E(\lambda_{min}) = 1/N$$

$$\mathrm{var}(\lambda_{min}) = (N - 1)/N^2(N + 1) \quad .$$

We might pause here to collect together the formulas giving the expected behavior of the eigenvalues, and their important functions, as a function of N, for $N \times N$ correlation matrices with random spectrum. Namely, we have seen that

$$E(\lambda_i) = 1 \qquad 1 \leqslant \lambda_i \leqslant N$$

$$E(\lambda_{max}) \approx \log N + \gamma$$

$$E(\lambda_{min}) = 1/N \text{ and}$$

$$E(\textstyle\sum \lambda_i^2) \approx 2N \quad .$$

Returning now to the joint distribution of $(\lambda_{max}, \lambda_{min})$, it can also be inferred from Reference 43 that these quantities are asymptotically independent, as a consequence of the formula

$$\Pr(\lambda_{min} > \frac{x}{N},\ \lambda_{max} < \log N - \log y)$$

$$\rightarrow \exp(-x - y),$$

as $N \rightarrow \infty$. This permits us to write, for large N,

$$E(\kappa(C)) = E(\lambda_{max}/\lambda_{min})$$

$$\approx E(\lambda_{max})\ E(1/\lambda_{min}).$$

However, although the first factor is finite, as we know, the second is not:

$$E(1/\lambda_{min}) = \int_0^1 \frac{1}{x}\ \frac{d}{dx}\left(1 - (1-x)^{N-1}\right) dx$$

$$= (N-1) \int_0^1 \frac{(1-x)^{N-1}}{x}\ dx$$

$$= (N-1) \int_0^1 \left(\frac{1}{x} + \ldots\right) dx = +\infty.$$

This observation suggests that the condition number $\kappa(C)$ may not have a finite first moment. Additional grounds for such suspicion can be based on its validity at the other extreme case where N = 2. In this simple case the assertion goes as follows: if a single number s is drawn at random from the interval [0, 1], and U (respectively V) is the min (respectively max) of {s, 1-s}, then the radio V/U obeys the distribution

$$\Pr(V/U \leq t) = \frac{t-1}{t+1},$$

and so has infinite expectation. This formula is derived by Feller (Reference 45, page 24). We now generalize this fact to the case of arbitrary N.

Theorem Let C be a correlation matrix with random spectrum. Then the condition number $\kappa(C)$ has infinite expectation.

Proof. In the notation of Section 4.1 we let $0 < u_{(1)} < u_{(2)} < \ldots < u_{(N-1)} < 1$ be the order statistics of a random sample of size N - 1 from the standard uniform distribution. The joint distribution P of these statistics is the ordered (N - 1)-variate Dirichlet distribution (Reference 35, Section 8.7), and is uniform over the region

$$\Omega = \{x: 0 < x_1 < x_2 < \ldots < x_{N-1} < 1\}$$

in R^N. Therefore,

$$E(\kappa(C)) = \int_{\Omega} \ldots \int \frac{\max\{u_{(1)},\ u_{(2)} - u_{(1)},\ldots,\ 1 - u_{(N-1)}\}}{\min\{\ \ldots\ldots\ \}}\, dP$$

$$= N(N-1)! \int_{T} \ldots \int \frac{\max\{\ \ldots\ \}}{u_{(1)}}\, du_{(1)} \ldots du_{(N-1)}$$

$$\geqslant (N-1)! \int_{T} \ldots \int \frac{1}{u_{(1)}}\, du_{(1)} \ldots du_{(N-1)}$$

where T is the subregion of Ω defined by

$$x_1 \leqslant \min\{x_2 - x_1,\ x_3 - x_2,\ldots,\ 1 - x_{N-1}\},$$

and we have used that the maximum spacing $\geqslant 1/N$. Now, the last multiple integral over T exceeds, for sufficiently small $\epsilon > 0$, the integrated integral

$$\int_0^{\epsilon} \frac{dx_1}{x_1} \int_{2x_1}^{1-(N-2)x_1} dx_2 \int_{x_1+x_2}^{1-(N-3)x_1} dx_3 \ \ldots \int_{x_1+x_{N-2}}^{1-x_1} dx_{N-1}$$

$$= \int_0^{\epsilon} \frac{1/(N-)! + x_1 q(x_1)}{x_1}\, dx_1$$

where q is a polynomial. This last integral is clearly divergent. QED

This completes our discussion of correlation matrices with random spectrum.

5. CORRELATION MATRICES WITH RANDOM GRAM STRUCTURE

In this final section we discuss random Gram matrices, as defined in Section 1.2, and briefly discussed in Section 2.2, along with some simulation results. We will follow the same plan as in the preceding section, namely, generating such matrices and distribution of certain related random variables. Finally we will make a few comparisons between the sample behavior of the two types of random matrices.

5.1 Method of Generation

We recall from Section 1.2 that an $N \times N$ random Gram matrix C has the form

$$C = TT^* \tag{5.1}$$

where the rows of T are i.i.d. vectors distributed uniformly on the sphere S^{N-1} in R^N. That is, for each row $\underline{t}_i$ of T, we have

$$\underline{t}_i \sim U(S^{N-1}) \quad .$$

So, just as in Section 4.1, the first question is how to express such random vectors in terms of standard univariate random variables.

This is a well researched problem, with contributions dating back at least 30 years. A short paper by Marsaglia[46] has a review of these early contributions, along with an improved method. More recent references are the pragmatic paper by Rubinstein,[47] and the extensive book of Devroye.[48] Again we distinguish between the analytic and the operational specification of $U(S^{N-1})$. The basic analytical result is that if X is a continuous radially symmetric N-dimensional random vector, then its projection on the sphere is uniformly distributed, that is,

$$X/\|X\| \sim U(S^{N-1}) \quad .$$

In particular, we can take $X \sim N(\theta, I)$, the standard spherical multivariate normal distribution. Operationally, the components of X can be generated by any of several standard pseudorandom normal variate routines. These eventually utilize pseudorandom uniform variates. The latter can also be used more directly to generate pseudorandom $U(S^{N-1})$ vectors, as is pointed out in (Reference 46 or in Reference 48, Chapter V). These are in addition to the brute force acceptance rejection method, which tends to be very inefficient for large N ($N \geq 5$, say). However, we will stick with the projected normally distributed random vectors.

Suppose now that we have a random vector $X \sim U(S^{N-1})$. It will be important to know how the components of X are distributed. It turns out that each

$$x_i^2 \sim \text{Beta}\left(\frac{1}{2}, \frac{N-1}{2}\right) \tag{5.2}$$

and that the density function of x_i is

$$C_N(1 - t^2)^{(N-3)/2} \quad , \qquad |t| < 1 \tag{5.3}$$

where $C_N = \Gamma((N/2)/(\Gamma(1/2)\Gamma((N-1)/2))$ is a normalizing constant. It is interesting to observe that these distributions vary considerably with dimension. In particular, we see that x_i follows an arc sine distribution when $(x_1,x_2) \sim U(S^1)$, while x_i is uniform on (-1,1) when $(x_1,x_2,x_3) \sim U(S^2)$. As N increases beyond 3, the density is unimodal with an ever steeper peak at t = 0.

We might note here that the joint density of two or more of the x_i is also available, as a consequence of some work of Stam.[49]

As a consequence of these facts we have the following geometrical lemmas: if X, Y are independently and uniformly distributed on S^{N-1}, then

$$E(<X,Y>) = 0 \tag{5.4a}$$

$$E(<X,Y>^2) = 1/N \tag{5.4b}$$

and

$$\text{var}(<X,Y>^2) = \frac{2(N-1)}{N^2(N+2)} \tag{5.4c}$$

Indeed, Equation (5.4a) is a consequence of the so-called "formula of total expectation,"

$$E(f(X,Y)) = E\,(E(f(X,Y)|X))$$

for scalar functions of two random vectors. The other two equations follow from realizing $<X,Y>^2$ as the squared length of the projection of a random point in S^{N-1} on a random axis, along with standard properties of the beta distribution. This geometrical information will be used below in the next two sections.

5.2 Distribution of Norms

As earlier, in Section 4.3, we will study the distributional behavior of $\|C\|_F$ and $\|C\|$, where C is now a random Gram matrix of the form of Equation (5.1), with the rows of T uniformly distributed over the unit sphere of appropriate dimension.

The study of $\|C\|_F^2$ is greatly facilitated by the preceding results, since these imply that the square of each off-diagonal entry of C has the beta distribution of Equation (5.2). It follows immediately that

$$E(\|C\|_F^2) = N + 2 \cdot \frac{1}{N} \cdot \frac{N(N-1)}{2}$$

$$= 2N - 1 \quad . \tag{5.5}$$

However, a variance formula is not so immediate, as we indicate next.

We consider the second moment of $\|C\|_F^2$ about the origin, that is,

$$E(\|C\|_F^4) = E\,((N + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} c_{ij}^2)^2). \tag{5.6}$$

Recalling that the first two moments of the beta distribution B(a,b) are $a/(a + b)$ and $a(a + 1)/(a + b)(a + b + 1)$, respectively, we have, upon expansion of Equation (5.6)

$$E(\|C\|_F^4) = N^2 + 4N \cdot \frac{1}{N} \cdot N(N - 1)/2$$

$$+ 4 \cdot \frac{3}{N(N + 2)} \cdot N(N - 1)/2 \qquad (5.7)$$

$$+ 4 \cdot N(N - 1)/2 \, [(N(N - 1)/2) - 1] \cdot x$$

where "x" is a generic notation for $E(c_{ij}^2 \, c_{kl}^2)$, for $i \neq k$ or $j \neq l$. Certainly if both $i \neq k$ and $j \neq l$, then $x = 1/N^2$, by independence.

In the remaining cases we are in the following situation. We have 3 random vectors u, v, w i.i.d. $U(S^{N-1})$ and we are considering the bivariate distribution of $(\langle u,v\rangle, \langle u,w\rangle)$. This distribution has also been considered by Stam,[49] who gave a formula for the density of the trivariate distribution of $(\langle u,v\rangle, \langle u,w\rangle, \langle v,w\rangle)$. He also proved that this distribution converges in total variation to the standard normal distribution on R^3. In view of the complicated nature of the aforementioned density, and of the rather rapid approach to the normal, as shown by simulations, we will ignore possible weak dependencies for small N, and use the approximation $x = 1/N^2$ throughout Equation (5.7). Therefore, after collecting terms there we arrive at the approximation

$$E(\|C\|_F^4) \approx 4N^2 - 4N - 1 + 6 \, \frac{N - 1}{N + 2} + \frac{2}{N} \quad . \qquad (5.8)$$

Simulations show this to be actually very accurate for $N \geq 5$. (In fact, extensive statistical testing never permits rejection of the hypothesis that the variates $\langle u,v\rangle$ and $\langle u,w\rangle$ are uncorrelated, for any N.) Finally we see that

$$\text{var}\,(\|C\|_F^2) \approx 6 \, \frac{N - 1}{N + 2} + \frac{2}{N} - 2$$

which, of course, is approximately 4 for large N.

These formulas for the first two moments of $\|C\|_F^2$ invite comparison with the corresponding formulas for correlation matrices with random spectra developed in the preceding section. While the means are very close, and asymptotically equivalent, there is a distinct difference in the behavior of the variances. Namely, the variance of $\|C\|_F^2$ when C has random spectrum varies as 4N, approximately, while that of $\|C\|_F^2$ when C is random Gram is asymptotically constant.

As earlier in Section 4.3 we expect that for fixed N, $\|C\|_F$ is approximately normal with mean $\sqrt{(2N - 1)}$, and this is borne out by simulations.

Writing $\|C\|_F^2$ in the form used in Equation (5.6), and appealing to the central limit theorem, the asymptotic normality follows readily:

$$\|C\|_F^2 \sim \text{Normal}\ (2N - 1,\ 4\frac{N^2 - 2N + 1}{N^2 + N})$$

for large N. As in the earlier section we could also establish the asymptotic normality of $\|C\|_F$, but at this point that can be left to the interested reader.

We now want to turn to the issue of the distribution of the spectral norm $\|C\|$ of a random Gram correlation matrix C. This particular topic brings us to the edge of a large and active field of research on the spectra of random matrices; see for instance Section II of the AMS Conference proceedings.[50] This area has a long history as indicated in the papers of Girko[51] and Geman[52] and their references, as well as the AMS volume. In turn it relates to many studies in the multivariate statistics field of spectral behavior of sample covariance matrices, see for instance T. Anderson's book.[1]

The essential observation runs as follows. We have $C = TT^*$ as defined in Equation (5.1). Then the columns of T^* are independent samples from the uniform distribution on S^{N-1}, and hence the matrix

$$S_N = \frac{1}{N} T^*T = \frac{1}{N} \sum_{k=1}^{N} t_k^* t_k$$

is the sample second moment matrix for such a distribution. (Here t_k is the kth row of the matrix T.) Now TT^* and T^*T always have the same eigenvalues, and hence, as a special case,

$$\|C\| = N\|S_N\| \quad . \tag{5.9}$$

With this observation we can now refer to the considerable body of previous work mentioned in the preceding references, and also to the book of Watson.[53] None of this seems to be exactly what we need. In particular, it is unlikely that we can ever know the precise distribution of $\|C\|$ for any fixed N. However, there are many asymptotic results. Here we will just take note of an improvement of Geman's theorem by Yin, Bai, and Krishnaiah, as referenced by Yin and Bai in Reference 50. Namely, let X_p be a $p \times n$ random matrix with i.i.d. entries, $n = n(p)$ an increasing function of p with

$$\lim_{p \to \infty} n(p) = y, \qquad 0 < y < \infty \quad .$$

Suppose that the entries of X_p have mean 0, variance σ^2, and finite fourth moment. Then

$$\lim_{p \to \infty} \|\frac{1}{n} X_p X_p^*\| = (1 + \sqrt{y})^2 \sigma^2 \tag{5.10}$$

almost surely.

We will now advance some reasons to support the conjecture that

$$\lim_{N \to \infty} \|C\| = 4 \tag{5.11}$$

almost surely when C is an $N \times N$ random Gram correlation matrix. First, from Equation (5.9) we have

$$\|C\| = N\|S_N\| = \|T^*T\|$$

$$= \| \frac{1}{N} X_p X_p^* \|$$

where $X_p = \sqrt{N}\, T^*$, $p = N$, and the columns of T^* are i.i.d. vectors $U(S^{N-1})$. Now the covariance matrix of any such random vector is $(1/N)I_N$, where I_N is the $N \times N$ identity matrix. [This can be seen by noting that the uniform distribution is invariant under unitary transformations, hence the covariance matrix commutes with all unitaries and so must be a multiple of I_N. That the multiple is $1/N$ follows from Equation (5.2).] Thus if the entries of each column were genuinely independent we would be in the situation where the limiting formula (5.10) applied, with $y = \sigma = 1$. In so far as this independence is present asymptotically, and that condition is sufficient, Equation (5.11) would follow.

A second approach is to begin with a random matrix $G = [g_{ij}]$, $1 \leqslant i, j \leqslant N$, with i.i.d. entries g_{ij}, each a standard normal variate. By Geman's theorem

$$\lim_{N \to \infty} \| \frac{1}{N} G\, G^* \| = 4,$$

almost surely. Now let D be the $N \times N$ diagonal matrix with i^{th} diagonal entry $= 1/\|i^{th}$ col. of $G\|$, and set $T^* = GD$. Then

$$\|C\| = N\, \|S_N\|$$

$$= \|T^*T\| = \|GD^2G^*\|$$

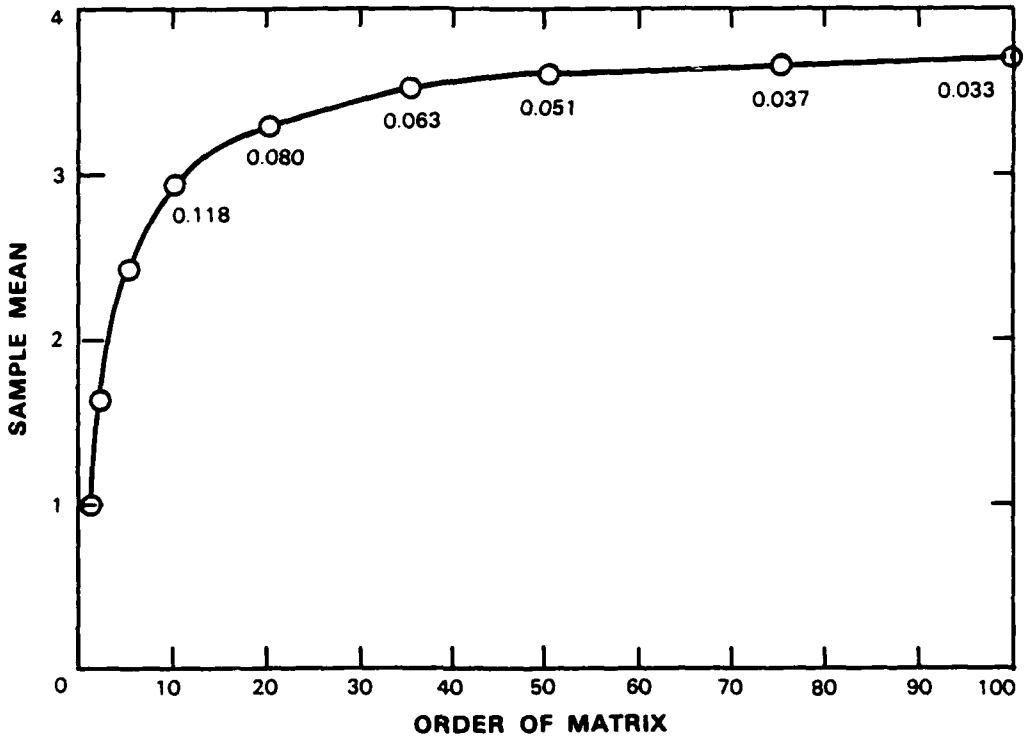
$$= \| \frac{1}{N} G(ND^2)\, G^* \|$$

and, in so far as $ND^2 \approx I_N$ for large N, we may expect Equation (5.11) to hold. This approach is, in effect, simply a more rigorous interpretation of the preceding approach, showing just how the lack of dependency down the various columns of C appears. Of course, there is a strong reason to believe that $ND^2 \to I_N$ 0 in some probabilistic sense, as $N \to \infty$, based both on simulation for $N \leqslant 200$, and on order statistic analysis beginning with the distribution of $\|i^{th}$ col. of $G\|^2$ as $chi^2(N)$.

Finally, we ran a brute force simulation for $N \leqslant 100$, and obtained the empirical curve of $E(\|C\|)$ against N shown in Figure 1. Each data point for $N \leqslant 50$ is based on 500 trials, while those for $N = 75$, 100 are based on 50 trials. The sample coefficient of variation is shown next to the data points.

5.3 Condition Number Expectation

In Section 4.3 it was shown that the condition number of a correlation matrix with random spectrum has an infinite first moment. In the present section we will demonstrate the analogous



101465-1

*Figure 1. Empirical spectral norm of C = TT**

fact for random correlation matrices of Gram type. The numerical results reported back in Section 2.2 (refer to Table 1), along with their theoretical result just mentioned, certainly have prepared us for this fact. Recall that the key empirical difference between the two main types of random correlation matrices was, in fact, the excessively ill-conditioned nature of random Gram matrices. We will discuss some other aspects of the spectral behavior of such matrices in the next section.

Theorem Let C be a random Gram correlation matrix. Then the condition number $\kappa(C)$ has infinite expectation.

Proof Making use of Taylor's inequality (3.15) we have C = TT* and

$$\kappa(C) \equiv \|C\| \, \|C^{-1}\| \geq 1/\min d_i^2$$

$$\geq 1/d_1^2$$

where $d_i \equiv \text{dist}(t_i, M_i)$, $\underline{t}_i$ = i^{th} row of T, so each t_i is i.i.d. $U(S^{N-1})$, and $M_i = \text{span}\{\underline{t}_j : j \neq i.\}$ Now since condim(M_i) = 1 in R^N, d_i is the magnitude of the projection of $\underline{t}_1$ on the line M_1. Let $\underline{u}_1$ be a unit vector in this subspace; then

$$d_1^2 = \langle \underline{t}_1, \underline{u}_1 \rangle^2 \quad .$$

That is, d_1^2 is the squared length of a random point on a random direction, and, as such, it has the distribution of Equation (5.2), with moments given by Equation (5.4). See also Reference 10. Therefore,

$$E\left(\kappa(C)\right) \geq E(1/d_1^2)$$

$$= \int_0^1 \frac{t^{-\frac{1}{2}} (1-t)^{\frac{N-3}{2}}}{B(\frac{1}{2}, \frac{N-1}{2})} \frac{dt}{t}$$

and this integral clearly diverges at 0. QED.

5.4 Empirical Spectral Behavior

This final section addresses the question "How random is the spectrum of a random Gram matrix?". Now, in one sense, this question has already been answered by the results of Sections 4.3 and 5.2. Namely, in those sections we derived the behavior of $\|C\|_F^2$, $\|C\|_F$, and $\|C\|$ for both types of random correlation matrices. Expressing these functions of C in terms of the eigenvalues shows that, not all spectral functions behave the same, and in particular, that random Gram matrices do not have a random spectrum. Below, we will briefly discuss some other aspects of this question.

Let us begin by considering the behavior of the smallest eigenvalue λ_N of an $N \times N$ random Gram matrix C. As in Section 5.2 there exists some related theoretical work in the literature (e.g., Reference 54), but again, it is not directly applicable to our situation. In view of the apparent boundedness of $\|C\|$ as $N \to \infty$, and of the simulation results from Section 2.2 which suggest that random Gram matrices are much more ill-conditioned than matrices with random spectrum, we might expect λ_N to be much smaller than the corresponding value from a correlation matrix with random spectrum. The distribution function of the latter is

$$F(x) = 1 - (1-x)^{N-1} \qquad 0 < x < 1$$

as follows from Equation (4.4), and we have

$$E(\lambda_{min})^2 = 1/N^2 \approx var(\lambda_{min})$$

as earlier noted.

Since we have, under the null hypothesis that the spectral behavior of C is random, the exact distribution and first two moments of λ_N, we can test this hypothesis by a variety of standard statistical tests. In particular, using the Kolmogorov-Smirnov one-sample test with 100 simulated random Gram matrices of various orders ($N \leq 20$), we are led to decisively reject the null hypothesis, at the 99 percent level.

Instead of considering the extreme eigenvalues λ_1, λ_N of C, we can inquire about the behavior, in some suitable sense, of the entire spectrum $\sigma(C)$. For example, we have already looked at the statistic

$$\|C\|_F^2 = \sum_1^N \lambda_i^2$$

and noted that asymptotically its mean behavior is that of a matrix with random spectrum, but its second moment behavior is quite different. Another approach is along the following lines. Beginning with the $N \times N$ random Gram matrix C, and its spectrum $\sigma(C) = \{\lambda_1, \ldots, \lambda_N\}$, we can invert the spacings method of Section 4.1 to arrive at a sample of points $\{x_1, \ldots, x_{N-1}\}$ in the unit interval. The transformation is

$$x_k = \frac{1}{N}(\lambda_{N-k+1} + \ldots + \lambda_{N-1} + \lambda_N)$$

for $1 \leqslant k \leqslant N - 1$. Under the null hypothesis this set of x's is a sample from the uniform distribution, and various statistics computed from this sample can be used for a test.

As an example, we considered Neyman's test[55] for uniformity. Fixing N, we generated a batch of 1000 random Gram matrices C, obtained their spectrum and the resulting points $\{x_1, \ldots, x_{N-1}\}$ in [0,1]. Then Neyman's statistic

$$N_2^2 = v_1^2 + v_2^2$$

was computed, where the v_j are the sample Fourier-Legendre coefficients when the density function f (from which the x's are drawn) is expanded in terms of Legendre polynomials:

$$f(x) = c \exp(1 + \sum c_j L_j(x)).$$

The motivation and theory of this test is discussed in the reference, and will not be given here. The distribution of N_2^2 is known approximately, and is asymptotically $chi^2(2)$. The null hypothesis is to be rejected for large values of N_2^2. For each N we calculated the fraction of the 1000 samples that exceeded various percentage points of the N_2^2 distribution, with the results indicated in Table 2. It is evident from these figures and the large number of trials that the null hypothesis of uniformity must be rejected. A closer examination of the data reveals that not only is there a very small eigenvalue λ_N, as noted above, but in fact there are enough small eigenvalues to pull the sample mean $\bar{x}$ far enough below 0.5 to greatly inflate the value of v_1 (precisely,

$$v_1 = \sqrt{(12n)}(x - \frac{1}{2})$$

where n = sample size = 1000, here). Incidentally, the sample coefficient of variation of the Neyman statistics decreased steadily from 0.27 at $N = 5$ to 0.045 at $N = 30$, showing very little scatter about the increasingly large values of N_2^2.

Finally, we offer two comments about the empirical behavior of the condition number of random Gram matrices. First, for various $N(\leqslant 20)$ we generated batches of 1000 each of random Gram matrices and correlation matrices with random spectra, and performed a Kolmogorov-Smirnov two-sample test on the respective condition numbers, to test the null hypothesis of a common distribution. This hypothesis was decisively rejected for all values of N, this rejection continued when the samples were subjected to trimming.

Second, bearing in mind the condition number bounds established in Section 3.2, we studied by simulation the tightness of the upper bound (3.18). That is, for various $N(\leq 50)$ we generated batches of random Gram matrices, computed their condition numbers, the co-linearity measure on the right hand side of Equation (3.15), and then their ratio as in Equation (3.18). The results are displayed in Table 3. They suggest that the admittedly crude upper bound in Equation (3.18) can indeed be reduced, and perhaps even be replaced by a term that is of order o(N).

TABLE 2

Fraction of Neyman Statistics Exceeding Various Percentage Points, and Sample Average

N (percent)	50	90	95	Mean N_2^2
5	99.8	45.	10.5	4.3
8	100.	98.5	84.4	7.0
10	—	100.	99.5	8.9
15	—	—	100.	13.3
20	—	—	—	17.7
30	—	—	—	26.5

TABLE 3

Empirical Ratio of Condition Number of Colinearity Measure for Random Gram Matrices

N	Batch Size	Sample Mean Ratio	Sample Coeff. of Var.
5	1000	5.20	.21
10	1000	8.82	.25
20	100	15.86	.24
35	100	22.65	.24
50	100	29.28	.22

6. SUMMARY

Let us summarize not only the foregoing technicalities, but also the place of this material in a larger scheme. In addition, we will point out several issues that remain to be resolved.

As noted at the outset, our interest in random correlation matrices stems from their interpretation as covariance matrices of purely random or 'average' (standardized) signals. A research project now underway has as its goal the evaluation of the efficacy of various group-theoretic signal processing algorithms. One ingredient that must be specified before a well-defined question can be posed in this context is a definite signal model. As remarked in the Introduction, such models can either be defined by a few (typically $\leqslant 2$) parameters, or they can be essentially nonparametric. A further possible subdivision of this latter class is into random stationary signals, or into purely random signals. The corresponding covariance matrices are then random correlation matrices with, in the first case, a Toeplitz structure. We have not discussed such special random matrices because it appears that, for practical purposes, most such behavior can be at least approximated by, for example, varying the parameters in an AR(2) signal model. Nevertheless, the question of generating random Toeplitz correlation matrices, and the statistical behavior of the corresponding entries, spectral functions, etc., is interesting, and is being studied, with results to be reported elsewhere.

We therefore have chosen to concentrate on random correlation matrices of the two principal types defined in Section 1.2, and studied in detail in Sections 2, 4, and 5. We observed early on that random Gram matrices exhibited a comparatively wild spectral behavior relative to correlation matrices with random spectrum. As we discovered later, this behavior is due to the presence, on average, of several much smaller eigenvalues than is consistent with the hypothesis of a random spectrum. In fact, a variety of both theoretical and empirical results shows that random Gram matrices do not have random spectrum; these are reviewed in Section 5.4.

In addition to collecting together numerous known results from the general statistics literature, and interpreting them in the present context of random correlation matrices, we have developed some new theoretical results. Specifically, in Section 3.2 we have extended the earlier work of J. Taylor on condition number lower bounds, and we assessed their tightness. This result is strictly deterministic. We then used this bound to show that the condition number of random Gram matrices (of a fixed size) has an infinite first moment. In view of our earlier empirical observations, this conclusion was not a complete surprise. Yet it also turned out that correlation matrices with random spectrum also have infinite first moment (for each fixed dimension $N \geqslant 2$, the case $N = 2$ being due to W. Feller).

We might offer an additional comment on the condition of random Gram matrices. Namely, referring back to the basic definition (Section 5.1), we could allow the row vectors t_i there to be drawn randomly from the unit sphere in a larger dimensional space. Geometric intuition suggests that with more 'room' in the sample space, collinearity should be less of a problem, with consequent improvement in conditioning. Numerical experiments show that, to an extent, this expectation is fulfilled. For example, in constrast with the data reported in Table 1, the mean (respective

median) condition number of 500 5×5 random Gram matrices based on vectors drawn uniformly from the sphere S^9 is 11.6 (respectively 8.7). The corresponding values for 500 10×10 random Gram matrices based on vectors drawn from S^{19} are 17.5 (respectively 15.2). However, it is not yet clear whether such higher dimensional random Gram matrices have random spectrum (eventually perhaps, but not initially!), or whether they have a finite first moment. This appears to be an attractive research area.

Finally, we note that some unresolved issues remain. In addition to specific technical questions, such as the best bound in Equation (3.18) (that is, the exact relation between the condition number of a Gram matrix and its Taylor lower bound), and validation of the conjectured limit of the spectral norm of large random Gram matrices in Equation (5.11), there is the issue of the intended application of these random correlation matrices to specific statistical testing procedures and simulations. In the present case the area of interest was described in the introduction. These matrices will serve to model random signals, the latter in turn serving as inputs to various signal processors defined by group filters and transforms, the objective being to assess the relative value of different groups of a common order (especially $N = 2^n$) for specific signal processing tasks such as Wiener filtering, decorrelation, data compression, etc. However, the question of whether one type of random correlation matrix should be preferred to another, for this particular applications, remains to be settled.

ACKNOWLEDGMENT

Mr. Thomas Loden very capably carried out the programming required for the numerical tests and simulations reported herein.

Mrs. Donna McTague was equally helpful in preparing the manuscript.

This work was sponsored by the Lincoln Laboratory Innovative Research Program (IRP).

REFERENCES

1. T. Anderson, *An Introduction to Multivariate Statistics*, 2nd ed. (Wiley, New York, 1984).
2. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, New York, 1972).
3. D. Kazakos, "Optimal Constrained Representation and Filtering of Signals," Signal Process, 347-353 (1983).
4. T. Anderson, I. Olkin, and L. Underhill, "Generation of Random Orthogonal Matrices," SIAM J. Sci. Statist. Comput. **8**, 625-629 (1987).
5. G. Stewart, "The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators," SIAM J. Numer. Anal. 17, 403-409 (1980).
6. L. Takacs, "Harmonic Analysis on Schur Algebras and Its Applications in the Theory of Probability," in *Probability Theory and Harmonic Analysis*, J-A. Chao and W. Woyczynski, eds. (Marcel Dekker, Inc., New York, 1986), 227-283.
7. P. Diaconis and M. Shahshahani, "The Subgroup Algorithm for Generating Uniform Random Variables," Prob. Eng. Inform. Sci. **1**, 15-32 (1987).
8. U. Grenander, *Probability on Algebraic Structures* (Wiley, New York, 1963).
9. L. Nachbin, *The Haar Integral* (van Nostrand, Princeton, 1965).
10. G. Marsaglia and I. Olkin, "Generating Correlation Matrices," SIAM. J. Sci. Statist., Comput. **5**, 470-475 (1984).
11. C. Chalmers, "Generation of Correlation Matrices with a Given Eigen-Structure," J. Stat. Comput. Simulation **4**, 133-139 (1975).
12. R. Bendel and R. Mickey, "Population Correlation Matrices for Sampling Experiments," Commun. Statist. Simul. Comput. B7, 163-182 (1978).
13. D. Johnson and W. Welch, "The Generation of Pseudo-Random Correlation Matrices," J. Stat. Comput. Simulation **11**, 55-69 (1980).
14. A. Horn, "Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix," Am. J. Math. **76**, 620-630 (1954).
15. L. Mirsky, "Matrices with Prescribed Characteristic Roots and Diagonal Elements," J. London Math. Soc. **33**, 14-21 (1958).
16. P. Fillmore, "On Similarity and the Diagonal of a Matrix," Am. Math. Monthly **76**, 167-168 (1969).
17. N. Chan and K. Li, "Diagonal Elements and Eigenvalues of a Real Symmetric Matrix," J. Math. Anal. Applic. **91**, 562-566 (1983).

18. J. Taylor, "The Condition of Gram Matrices and Related Problems," Proc. Royal Soc. Edin. 90A, 45-56 (1978).

19. G. Stewart, "Collinearity and Least Squares Regression," Stat. Sci. **2**, 68-83 (1987).

20. R. Horn and C. Johnson, *Matrix Analysis* (Cambridge, Univ. Press. New York, 1985).

21. P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. (Academic Press, New York, 1985).

22. G. Stewart, *Introduction to Matrix Computations* (Academic Press, New York, 1973).

23. G. Golub and C. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1983).

24. R. Holmes, "A Formula for the Spectral Radius of an Operator," Am. Math. Monthly **75**, 163-166 (1968).

25. A. Leclerc, "Une borne superieure pour les valeurs d'une matrice symetrique. Applications," C.R. Acad. Sci. Paris **287**, A553-555 (1978).

26. T. Kato, "Estimation of Iterated Matrices, with Application to the von Neumann Condition," Numer. Math. **26**, 22-29 (1960).

27. E. Beckenbach and R. Bellman, *Inequalities* (Springer-Verlag, Berlin, 1961).

28. J. Riley, "Solving Systems of Linear Equations with a Positive Definite, Symmetric, but Possibly Ill-Conditioned Matrix," Math. Computation **9**, 56-61 (1955).

29. A. Marshall and I. Olkin, "Norms and Inequalities for Condition Numbers, II" Linear Alg. Appl. **2**, 167-172 (1969).

30. A. Cline et. al., "An Estimate for the Condition Number of a Matrix," SIAM J. Numer. Anal. **16**, 368-375 (1979).

31. J. Dixon, "Estimating Extremal Eigenvalues and Condition Numbers of Matrices," SIAM J. Numer Anal. **20**, 812-814 (1983).

32. W. Hager, "Condition Estimates," SIAM J. Sci. Statist. Comput. **5**, 311-316 (1984).

33. W. Greub and W. Rheinboldt, "On a Generalization of an Inequality of L.V. Kantorovich," Proc. Am. Math. Soc. **10**, 407-415 (1959).

34. G. Styan, "Hadamard Products and Multivariate Statistical Analysis," Linear Alg. Appl. **6**, 217-240 (1973).

35. S. Wils, *Mathematical Statistics* (Wiley, New York, 1962).

36. A. Siegel, "Testing for Periodicity in a Time Series," J. Am. Stat. Assoc. **75**, 345-348 (1980).

37. R. Pyke, "Spacings," J. Royal Stat. Soc. **B27**, 395-436 (1965).

38. P. Deheuvels, "Spacings and Applications," Proc. 4th Pannonian Symp. on Math. Stat., Austria 1983, 1-30.

39. F. Steutel, "Random Division of an Interval," Stat. Neerlandica **21**, 231-244 (1967).

40. M. Kendall and P. Moran, *Geometrical Probability* (Griffin, London, 1963).

41. M. Greenwood, "The Statistical Study of Infectious Diseases," J.R. Stat. Soc. London A **109**, 85-110 (1946).

42. P. Moran, "The Random Division of an Interval," Proc. Cambridge Philos. Soc. **B9**, 92-98 (1947).

43. D. Darling, "On a Class of Problems Related to the Random Division of an Interval," Ann. Math. Stat. **24**, 239-253 (1953).

44. L. Devroye, "Laws of the Iterated Algorithm for Order Statistics of Uniform Spacings," Ann. Prob. **9**, 860-867 (1981).

45. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2 (Wiley, New York, 1971).

46. G. Marsaglia, "Choosing a Point from the Surface of a Sphere," Ann. Math. Stat. **43**, 645-646 (1972).

47. R. Rubinstein, "Generating Random Vectors Uniformly Distributed Inside and on the Surface of Different Regions," Eur. J. Oper. Res. **10**, 205-209 (1982).

48. L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 1986).

49. A. Stam, "Limit Theorems for Uniform Distributions on Spheres in High-Dimensional Euclidean Spaces," J. Appl. Probability **19**, 221-228 (1982).

50. "Random Matrices and their Applications," Contemporary Math. **50** (Am. Math. Soc., Providence, 1986).

51. V. Girko, "Spectral Theory of Random Matrices," Russ. Math. Surv. **40**, 77-120 (1984).

52. S. Geman, "A Limit Theorem for the Norm of Random Matrices," Ann. Prob. **8**, 252-261 (1980).

53. G. Watson, *Statistics on Spheres* (Wiley, New York, 1983).

54. J. Silverstein, "The Smallest Eigenvalue of a Large Dimensional Wishart Matrix," Ann. Prob. **13**, 1364-1368 (1985).

55. H. Solomon and M. Stephens, "On Neyman's Statistic for Testing Uniformity," Commun. Statist. **B12**, 127-134 (1983).'

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report 803	5. MONITORING ORGANIZATION REPORT NUMBER(S) ESD-TR-88-158
6a. NAME OF PERFORMING ORGANIZATION Lincoln Laboratory, MIT 6b. OFFICE SYMBOL *(If applicable)*	7a. NAME OF MONITORING ORGANIZATION Electronic Systems Division
6c. ADDRESS *(City, State, and Zip Code)* P.O. Box 73 Lexington, MA 02173-0073	7b. ADDRESS *(City, State, and Zip Code)* Hanscom AFB, MA 01731
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Systems Command, USAF 8b. OFFICE SYMBOL *(If applicable)*	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-85-C-0002
8c. ADDRESS *(City, State, and Zip Code)* Andrews AFB Washington, DC 20334	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. 359 / TASK NO. / WORK UNIT ACCESSION NO.
11. TITLE *(Include Security Classification)* On Random Correlation Matrices	
12. PERSONAL AUTHOR(S) Richard B. Holmes	
13a. TYPE OF REPORT Technical Report 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT *(Year, Month, Day)* 28 October 1988 15. PAGE COUNT 60
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	correlation matrix, random Gram matrix, spacings, random correlation matrix, condition number, Neyman's test, random spectrum, random eigenvalue

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report contains a detailed study of random correlation matrices, including algebraic, statistical, and historical background. Such matrices are of particular interest because they serve to model 'average signals' for simulation testing of signal processing algorithms. The latter half of this report extensively discusses the statistical behavior of spectral functions of the two major types of random correlation matrices, from both theoretical and empirical apsects. Our emphasis then, is on eigenvalue distribution and condition number behavior. Actual application to algorithm testing will be described in a subsequent report.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Lt. Col. Hugh L. Southall, USAF	22b. TELEPHONE *(Include Area Code)* (617) 981-2330 22c. OFFICE SYMBOL ESD/TML

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE